The pitch hunt

Clarke, Jeanne Nora

Publication date:
2017

Chapter 3

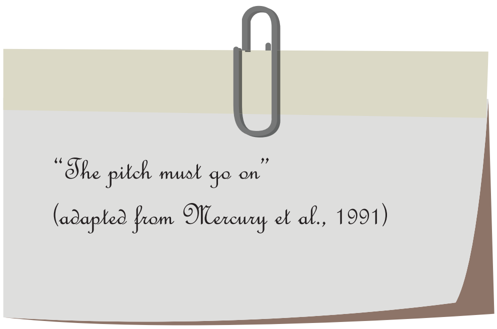
Effect of F0 contours on top-down repair of interrupted speech: Ain't that a pitch!

Clarke, J.,
Baskent, D.,
Gaudrain, E.

A shorter version of this chapter has been submitted to JASA-EL

Abstract

Top-down repair of interrupted speech can be influenced by bottom-up acoustic cues such as voice pitch (F0). This study aims to investigate the role of the dynamic information of pitch, i.e. F0 contours, in top-down repair of speech. Intelligibility of sentences interrupted with silence or noise was measured in five F0 contours conditions (original, flat, exaggerated with a factor 1.5 and 1.75, inverted). Our main hypothesis is that natural F0 contours would better guide the listener to successfully link successive segments of interrupted speech, a necessary step to restore interrupted speech, and that manipulating F0 contours would impair this linking and thus negatively affect top-down repair. Intelligibility of interrupted speech was impaired only by misleading dynamic information (inverted F0 contours). The top-down repair of interrupted speech was not affected by any F0 contours manipulation.



3.1. Introduction

Speech perception can be challenging in the presence of noisy background. The phonemic restoration paradigm (Warren, 1970) can be used to study the brain's ability to reconstruct periodically interrupted sentences. Besides linguistic knowledge, expectations and context (Bashford et al., 1992; Samuel, 1981; Verschuure and Brocaar, 1983), top-down repair of speech is also affected by the availability of acoustic bottom-up cues (Bhargava et al., 2014; Clarke et al., 2016). Voice pitch, the perceptual correlate of the fundamental frequency (F0), has been identified as an important bottom-up cue to help speech in noise perception. It is a strong across-frequency grouping cue (Darwin and Carlyon, 1995), i.e. (i) pitch information brings coherence to speech sounds by fusing together different parts of the spectrum which help phoneme identification, and (ii) average F0 and/or dynamic patterns (F0 contours) help to segregate different sound sources, which is useful to attend to target speech in the presence of maskers. Both mechanisms directly affect intelligibility of speech segments. The present study, in line with previous studies from our lab, investigates whether voice pitch is also used for sequentially linking successive speech segments across interruptions. In interrupted speech, we previously showed that alternating the average F0 value did not hinder intelligibility of interrupted speech nor phonemic restoration benefit (Clarke et al., 2014). This suggests that the continuity of the average F0 value might not be necessary for linking successive speech segments across a noise gap. However, the total absence of F0 (unvoiced speech) hindered intelligibility of interrupted speech compared to correct pattern of voicing (normal speech), whereas the addition of noise in silent gaps still improved intelligibility, thus showing a restoration benefit (Clarke et al., 2016). This suggests that even without F0 cues, unvoiced speech can still be segregated from noise. This would help the noise to hide the spurious cues introduced by the sudden silent gaps, but, on the other hand, also still allow the unvoiced speech segments to be linked across the noise bursts.

The goal of this study is to investigate the effect of the magnitude and the direction of F0 contours on intelligibility of interrupted speech (with silence and noise) and on phonemic restoration. We used similar F0 contour manipulations as previous studies that consistently showed an effect of modifying F0 contours for speech perception (Binns and Culling, 2007; Meister et al., 2011; Miller et al., 2010). In the present study,

3

the modifications of F0 contours consisted of (i) misrepresenting the dynamic F0 information by inverting the F0 contours within the same magnitude (inverted F0), (ii) compressing the magnitude of F0 around its median value, thus removing the dynamic information of the F0 contours (flat F0), (iii) expanding the magnitude of F0 by exaggerating the F0 contours with a factor 1.5 (exaggerated 1.5), and (iv) with a factor 1.75 (exaggerated 1.75). Compared to the original F0 contours (original F0), the inverted F0 condition would allow to investigate the effect of misleading F0 cues (e.g. for accentuations and surrounding contour guiding to content words, Binns and Culling, 2007) independently of F0 magnitude, that is itself investigated in the other conditions.

First, the F0 contour manipulations can affect intelligibility of interrupted speech at the speech segment level. Reducing F0 contour's magnitude (flat F0 contour) and changing the direction of F0 contours (inverted F0 contour) are expected to impair phoneme identification and coarticulation, because across frequency grouping with these manipulated F0 contours might produce more errors. Thus flat and inverted F0 contour might have a negative effect on intelligibility at the speech segments level. However, with exaggerated F0 contours, phonemes are more contrasted and provided they still correspond to proper categories, perception of the independent speech segments could be facilitated (such as in infant-directed speech - Kuhl et al., 1997).

Second, F0 seems to contribute to a robust linking of speech segments, but likely not from average F0 (Clarke et al., 2014). Therefore, here, we hypothesize that its dynamic fluctuations (F0 contours) are used for linking successive speech segments. Specifically, we expect natural F0 contours to better guide the listener to successfully link successive segments of interrupted speech. The predictive nature of F0 contours supports this hypothesis. A simple example is from interrupted tone glides that are perceived continuous behind the masking noise (Dannenbring, 1976). This suggests that the dynamic information helps to predict shift in frequency of the tone glide after the noise interruption and that it is restored behind the masking noise. Similarly, the dynamic fluctuations of F0 in interrupted sentences may play a role in top-down repair of speech. We tested different F0 contour patterns (inverted, flat, original, and exaggerated contours with two ratios). As already mentioned, F0 contours have been shown to be important for stream segregation (e.g. for speech in noise: Meister et al., 2011; Miller et al., 2010; and even more so for competing talkers: Binns and Culling, 2007; Wang et al., 2013), with misleading (inverted) contours eliciting worse performance than no (flat) or expanded (exaggerated) contours. Thus, we expected inverted contours to impair linking successive speech segments the most because of the misleading dynamic information it provides, whereas flat contours, without dynamic F0 information, would not help nor impair linking successive speech segments. For the exaggerated F0 contours, one might hypothesize that larger frequency jumps (i.e. greater manipulating ratio) at the interruptions would be more difficult to link, but Clarke et al. (2014) showed that interrupted speech intelligibility and phonemic restoration were not affected by F0 alternation of octave between successive speech segments. Thus we were not expecting exaggerated F0 contour to impair linking successive speech segments. However, F0 contours have another function at the sentential level besides linking speech segments across time that might affect interrupted speech performance and top-down repair mechanisms.

Besides F0 tracking to link speech segments across interruptions, F0 contours, as a primary feature contributing to prosody, also gives information on the intonation of an utterance. Some linguistic functions that can be associated with F0 contours are segmentation (word boundaries) and lexical stress (used for segmentation in English and in Dutch), accentuation (focus on important words in a sentence), and types of utterance (statement or question), as well as lexical meaning in tonal languages (Cutler et al., 1997; Cutler and Donselaar, 2001; Eady and Cooper, 1986; Gussenhoven, 2004; Liu and Rodriguez, 2012; Quam and Swingley, 2014; Spitzer et al., 2007; Wang et al., 2013). All this can contribute to the overall intelligibility of interrupted sentences. However, Chatterjee et al. (2010) compared flat with original F0 contour in speech interrupted with silence, and showed that there was no effect of the intonation contour on intelligibility. They suggested that sentence context could partially compensate for the missing dynamic information with flat F0 contour in silence. However, it is also possible that without competing signal (as is the case with silent interruptions), pitch is not a necessary linking cue, as speech contains redundant information (Assmann and Summerfield, 2004). With information transmitted in different forms in speech, redundancy facilitates decoding of information especially when there are losses in the transmission (e.g. due to noise or interruptions). In our study, we expected inverted F0 contour to have a detrimental effect on intelligibility at the sentential level because lexical stress and accentuation would be misleading.

Instead of highlighting important information, inverted F0 contour would have the opposite effect, i.e. the important information that is usually accentuated with normal F0 contour would be attenuated by the inversion of F0 contour. Moreover, in line with Chatterjee and colleagues (2010), we expected no effect of flat F0 contour on intelligibility at the sentential level, as the lack of dynamic information could be compensated by linguistic context. Furthermore, it is possible that exaggerated F0 contour might strengthen lexical stress and accentuation, with important information even more highlighted (proportionally with the expansion ratio). We can thus expect a positive effect of exaggerated F0 contour on intelligibility at the sentential level. At the sentential level, the range of F0 contour manipulations covers different expectations on intelligibility performance.

Overall, taking into account the expectations at the different levels, i.e. individual speech segments intelligibility, linking of successive speech segments, and intelligibility of the whole sentence, we expected to have better performance for the exaggerated contour, the flat contour, and finally for the inverted contours.

3.2. Methods

Participants

Sixteen native Dutch speakers with normal hearing (20 dB HL or less pure-tone thresholds at audiometric frequencies of 250-6000 Hz in both ears), aged between 20 and 40 years (mean=25.5, s.d.=6.2), and with no hearing or speech-related problems (self-reported), participated in this study. The study was approved by the Medisch Etische Toetsingscommissie (Medical Ethical Review Committee) of the University Medical Center Groningen. All participants were informed about the procedure and signed a consent form. Participants were paid for their participation.

Stimuli

The speech stimuli of this study were Dutch sentences with high sentential-context, spoken by a male talker and digitized at 44.1 kHz sampling rate (Versfeld et al., 2000). Each sentence is grammatically and syntactically correct. The sentences consist of four to nine words and each word is maximum three syllables. The corpus is divided into 39 homogeneous subsets of 13 sentences. Each sentence in the subsets is equally intelligible. Each subset has the same distribution of phonemes with the average frequency of phonemes in Dutch. In this study, twenty subsets of the corpus were used (5 for baseline, 5 for familiarization and 10 from set numbers 14 to 23 for data collection).

Signal processing

The fundamental frequency (F0) contours were manipulated offline using TANDEM-STRAIGHT in MATLAB (Kawahara and Morise, 2011). The sentences used in the experiment were processed under 5 F0 contour conditions (see Figure 3.1): inverted F0, flat F0, original F0, exaggerated F0 by a factor of 1.5, and exaggerated F0 by a factor of 1.75. For all manipulations, only the voiced segments of the sentences were manipulated. The F0 contours were manipulated with the following formula, similar to the procedure described in Binns and Culling (2007), Miller et al. (2010) and Meister et al. (2011):

$$F_0' = F_{0med} \times \left(\frac{F_0}{F_{0med}}\right)^{\alpha} \quad (1)$$

3

where F_0 is the original F0, $F_{0_{med}}$ is the median F0, $F_0{}'$ is the manipulated F0, and α is the modification ratio ($\alpha = 0$, for flat contours ; $\alpha = -1$, for inverted contours ; $\alpha > 1$, for exaggerated contours). Note that equation 1 is valid for linear frequencies.

For the inverted F0 condition, the symmetry about the median F0 value was used in the logarithmic scale. For the flat F0 condition, the median F0 replaced the original F0. For the exaggerated F0 by a factor 1.5 and 1.75, the F0 contours were expanded in the logarithmic scale by factor 1.5 and 1.75.

The resynthesized sentences were modulated with a square wave to obtain an interrupted sentence, with an interruption rate of 2.2 Hz, and a duty cycle of 50%; same interruption parameters as Clarke et al. (2014). A raised cosine ramp of 5 ms was applied to the onsets and offsets of the square wave to reduce spectral splatter. Silent and speech shaped noise (SNR of -5dB) were used for interruption. For each condition, a separate filler noise file was generated with white noise modulated by the long-term average spectrum of all sentences in the given condition (as done in Clarke et al., 2014).

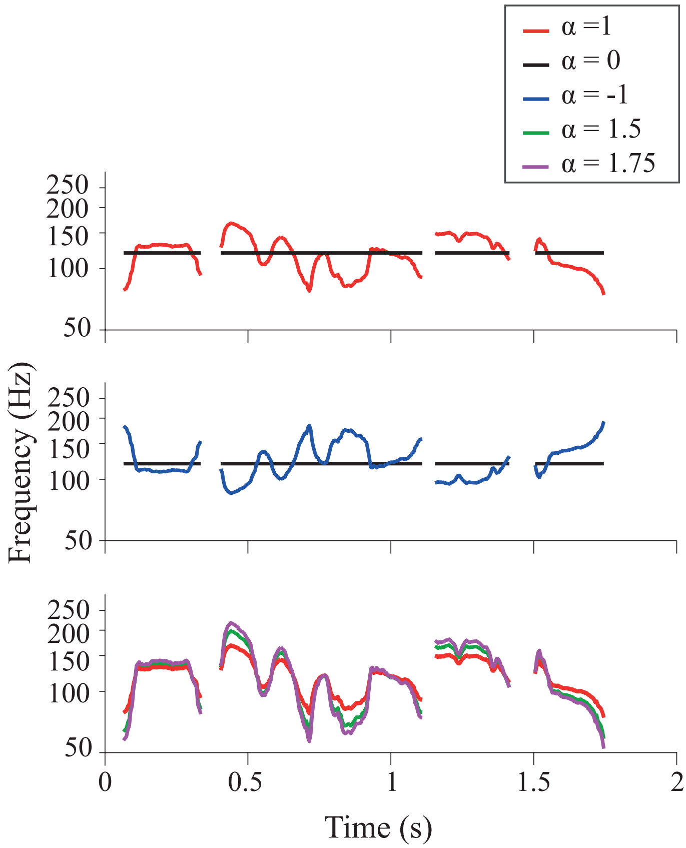


Figure 3.1. Manipulated F0 contours for the sentence "Buiten is het donker en koud" (Outside it is dark and cold). The original F0 contour (α = 1) is in red, the flattened F0 (α = 0) is in black, the inverted F0 (α = -1) is in blue, the exaggerated F0 by a factor 1.5 (α = 1.5) and 1.75 (α = 1.75) are in green and magenta, respectively.

Apparatus

The participants were seated in a sound-attenuated booth during the experiment. The stimuli were sent through the S/PDIF output of an AudioFire 4 soundcard (Echo Digital Audio Corporation). They were converted to an analog signal via a DA10 D/A converter (Lavry Engineering Inc.). The participants listened to the sentences diotically through HD600 headphones (Sennheiser Electronic Corporation) at an RMS level of 65 dB SPL. We used a Palm Track digital voice recorder (ALESIS) and microphone (SHURE PG48) to record the responses of the participants for offline scoring by native Dutch student assistants.

Procedure

Participants came for a single session which lasted around an hour and a half and which included obtaining written informed consents, conducting the audiometric test, the baseline measurement, familiarization and data collection, the debriefing, and occasional breaks. To measure the baseline, the first five lists of the sentences were used without interruption. Each list was presented with a different F0 contour condition, and all lists and conditions were randomized for each participant. Participants then familiarized with the procedure listening to five conditions selected randomly out of the ten conditions of the experiment. The familiarization phase was similar to the experiment expect that written and auditory feedback were provided. After responding, the participants could read the complete sentence as they listened to the unprocessed version of the sentence followed by the interrupted sentence once again.

The data collection consisted of 10 conditions: 5 F0 contour conditions (inverted F0 contour, flat F0 contour, original F0 contour, exaggerated F0 contour by 1.5 and 1.75) × 2 interruption conditions (silent intervals and filler noise). Both the sentence lists and the conditions were presented in random order. At the beginning of each set the participants heard the same introduction sentence to prepare them for the trial condition. A tone preceded each sentence to alert the participant. After hearing the sentence, the participants were asked to repeat what they could understand from the sentence stimuli, and were additionally encouraged to guess as much as possible. Each word in the sentence was scored according to participants' correct response. Total rationalized arcsine transformed unit (RAU) scores were computed for each condition (Studebaker, 1985).

3.3. Results

The upper panel of Figure 3.2 displays the intelligibility scores in each of the ten conditions (5 F0 contours x 2 interruptions). A repeated measure two-way analysis of variance (ANOVA) was performed on the RAU scores with F0 contours (5 levels) and interruption (2 levels) as the within-subject factors. The effect size is indicated by eta square, η^2 (Bakeman, 2005). A significant effect of F0 contours [$F(4,60) = 4.20$, $p = 0.0046$, $\eta^2 = 0.063$] indicated that F0 contours have an overall effect on intelligibility of interrupted speech. A significant effect of interruption [$F(1,15) = 71.63$, $p < 0.001$, $\eta^2 = 0.15$] indicated the presence of phonemic restoration effect. However, there was no interaction between the two factors [$F(4,60) = 0.16$, $p = 0.95$, $\eta^2 = 0.0036$] which indicated that the F0 contour manipulations had the same effect on intelligibility with silent and noise interruptions, i.e. that the manipulations of the F0 contours did not affect phonemic restoration. The results from the phonemic restoration for the manipulated F0 contours are displayed in the lower panel of Figure 3.2. Phonemic restoration effect was computed by subtracting the scores in the silent condition from those in the noise condition for each F0 contour condition.

The overall effect of F0 contours on intelligibility was small, as shown by the small effect size ($\eta^2 = 0.063$) that indicated that only 6 % of the variance of intelligibility was explained by the F0 contour manipulations. Differences between only some pairs of conditions were observed. Table 3.1 shows the outcome of a post-hoc analysis using pair-wised t-tests with FDR control conducted to compare performance of F0 contours between each other (averaged on noise and silent conditions, because of the lack of interaction between the two factors). The results showed that intelligibility performance with inverted F0 contour was significantly poorer from that of other contours except that of the flat F0 contour. All other comparisons were not significantly different.

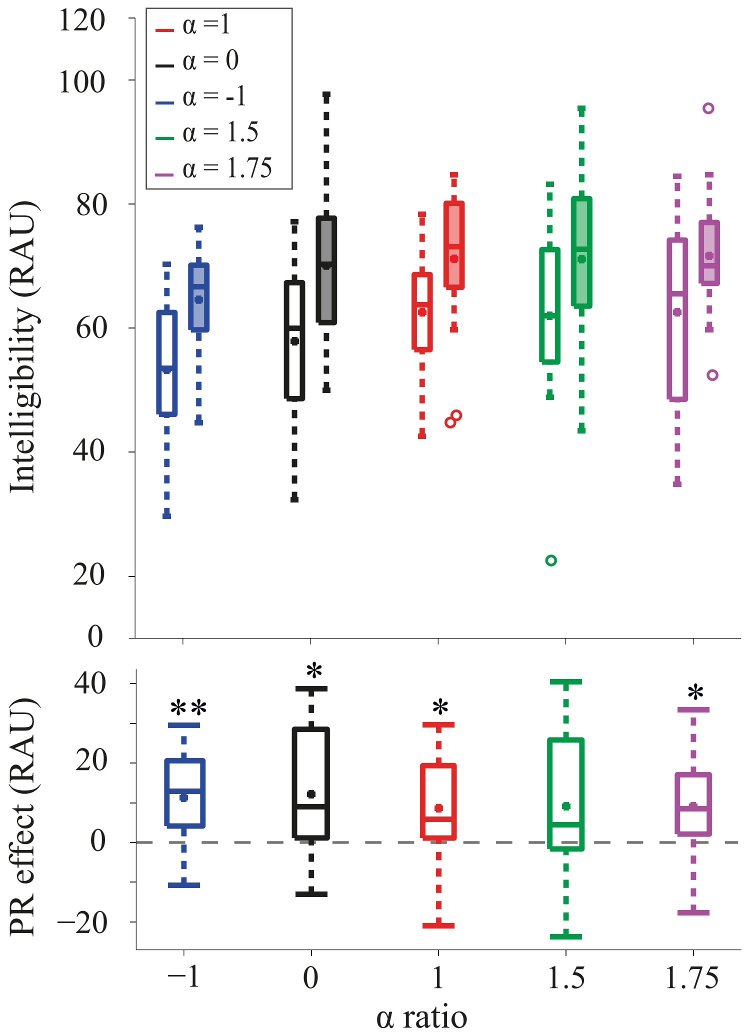


Figure 3.2. Intelligibility results (top panel) for silent interruptions (empty boxes) and noise interruptions (filled boxes) and phonemic restoration effect (lower panel) for the manipulated F0 contours, from left to right: inverted (α = -1), flat (α = 0), original (α = 1), exaggerated with a factor 1.5 (α = 1.5), and 1.75 (α = 1.75). The horizontal line indicates the median, the box indicates the 25th and 75th quartiles, and the dashed whiskers indicate the 1.5 interquartile range (IRQ). The circles indicate the outliers. The dots indicate the mean.

Table 3.1. Comparisons of overall intelligibility for each F0 contour conditions. Significant p-values (adjusted with FDR) are highlighted in grey cells.

	α = -1	**α = 0**	**α = 1**	**α = 1.5**
α = 0	t= 2.0075 p = 0.13			
α = 1	t= 4.023 p = 0.0034	t= 1.15 p = 0.43		
α = 1.5	t= 2.84 p = 0.026	t= 0.85 p = 0.57	t= -0.14 p = 0.93	
α = 1.75	t= 3.60 p = 0.0054	t= 1.15 p = 0.43	t= 0.085 p = 0.93	t= 0.18 p = 0.93

There was no effect of F0 contours on the phonemic restoration benefit as indicated by the lack of interaction in the ANOVA on the intelligibility scores. However, as an a priori variable, we tested whether the phonemic restoration scores were significantly different from 0 with a one-sample t-test for each F0 contour condition (displayed in Table 3.2 and indicated by a black star on the lower panel of Figure 3.2). A significant phonemic restoration benefit was observed in all F0 contour conditions except one, the exaggerated F0 contour with ratio 1.5. This indicates that F0 contour manipulations did not affect top-down repair mechanisms of interrupted speech.

Table 3.2. Comparisons of PR scores to 0 for each F0 contour condition. Significant p-values are highlighted in grey cells.

	α = -1	**α = 0**	**α = 1**	**α = 1.5**	**α = 1.75**
PR ≠ 0	t = 3.87 p = 0.0015	t = 2.90 p = 0.011	t = 2.53 p = 0.023	t = 2.022 p = 0.061	t = 2.88 p = 0.012

3.4. Discussion

We were interested in investigating the effects of the magnitude and direction of F0 contours on intelligibility and top-down repair of interrupted speech. We expected that both magnitude and direction manipulations of F0 would affect comprehension of isolated speech segments as well as linking successive speech segments into a coherent speech stream, and thus produce a range of performance of global intelligibility of interrupted speech as well as the top-down repair. However, we showed that modifying the magnitude of F0 contours (all conditions except the inverted contours) did not have any effect on interrupted speech intelligibility (in line with Chatterjee et al., 2010 for flat F0 contour). On the other hand, the direction of the F0 contours seems to be a cue listeners relied on for intelligibility of interrupted speech. Indeed, partially validating our hypothesis, having misleading dynamic information of F0 (inverted contours) impaired interrupted speech intelligibility. Nevertheless, even the misleading dynamic information of F0 did not impair the top-down repair of interrupted speech. This result suggests that participants may have compensated for the atypical F0 cues with the linguistic context (as suggested by Chatterjee et al., 2010). In this study, it seems that top-down repair of speech may rely more on linguistic cues than on F0 cues (Clarke et al., 2014).

For the global intelligibility scores, inverted F0 contour, which provides misleading intonation cues, were the only manipulated F0 contours to show a decrease of performance from the original F0 contour. This confirms that wrong F0 dynamic information leads to lower intelligibility of interrupted speech, as was found for speech with continuous background interferer (Binns and Culling, 2007; Meister et al., 2011; Miller et al., 2010). An explanation can be that original F0 contour help to define clause boundaries whereas inverted F0 contour distort those boundaries. Indeed, Wingfield et al. (1984) showed that listeners better report a complete syntactic unit (such as a sentence) than an incomplete clause. In a sentence, the F0 contours have expected patterns, such as falling at the end of a statement or rising at the end of a closed question, whereas an incomplete clause can disturb the listener’s expectations, impairing sentence intelligibility. However, our other F0 contour manipulations did not show any difference in performance, contrary to other studies with continuous background interferer (Binns and Culling, 2007; Laures and Bunton, 2003; Miller et al., 2010; Wang et al., 2013). Nevertheless, as already pointed out for interrupted speech with silence, Chatterjee et al. (2010) did not show reduced intelligibility with flat F0 contour. Finding a similar lack of effect of flattening F0, even with a different interruption rate (5Hz in Chatterjee et al.'s study against 1.5Hz in the present study), seems to confirm that removing F0 dynamic information does not hinder intelligibility of interrupted speech segments. This may also suggest that lack of intonation patterns can be overcome for interrupted speech perception, suggesting again the prevalence of linguistic cues. For the exaggerated F0 contour, we used two expansion ratios (i.e., 1.5 and 1.75) to investigate the effect of the F0 variations magnitude. The two exaggerated F0 contour conditions did not yield any performance differences from each other, nor with the other conditions (except the inverted F0 contour). These results do not confirm our hypothesis on better phonetic categorization improving speech perception with wider F0 variations, and also confirm our hypothesis that wider F0 variations do not weaken linking successive speech segments (in line with Clarke et al. 2014). Taken all together, these results suggest that the F0 variations' magnitude may not be used as a linking cue for interrupted speech perception, but that direction of F0 contours did. Moreover, only when all three aspects of interrupted speech perception (individual speech segments intelligibility, linking successive speech segments, and sentential intelligibility) were affected by the F0 contour manipulation, ie. for the inverted F0 contour, did overall intelligibility decrease. This suggests that participants seem to fail to compensate for the inverted F0 contour manipulation that impairs more aspects of interrupted speech perception than our other F0 contour manipulations.

Second, a significant phonemic restoration benefit was observed for all F0 contour conditions, indicating that participants performed better when the silent interruptions were filled with noise bursts. This result is unexpected for the inverted F0 contour, as we predicted that the atypical F0 contours would disturb following the speech stream (as well as impair speech perception — confirmed by decreased intelligibility scores for interrupted speech). However, it seems that the weaker linking of successive speech segments hindered by the misleading F0 dynamic information could be compensated for. This suggests that speech redundancy was sufficient to still perform top-down repair of speech and that other cues, such as the linguistic context, might have helped to overcome the misleading F0 cues for building up a coherent speech stream. But as intelligibility was affected by the inverted F0 contour, it suggests that misleading F0 dynamic cues impacted some other linguistic information such as accentuation (lexical stress) and access to content word (sentence stress). For the flat F0 contour, we expected no effect of the lack of F0 variations on linking successive speech segments. Thus, we expected the restoration benefit to remain, as observed. As intelligibility was also not affected by flattening the F0 contours, this suggests that F0 variations are not necessary for both intelligibility and top-down repair of interrupted speech, likely because of the prevalence of the linguistic contextual cues in the present task. However, F0 variations may be necessary for other tasks, such as recognizing emotions, and might thus be more difficult to compensate for in such a case. Exaggerated F0 contour (with both expansion ratios) had no effect on top-down repair of speech or on intelligibility, again showing that sequentially linking successive speech segments was not affected by F0 contours magnitude. It suggests that the noise bursts filling the silent interruptions still acted as a masker when F0 contours were exaggerated, indicating that linking speech segments with wider F0 variations across the noise bursts may have been perceptually possible, in line with results from Clarke et al. (2014). The present study suggests that the F0 contour manipulations, weakening the sequential linking of successive speech segments (for inverted F0 contour), did not affect phonemic restoration benefit. Thus, for interrupted speech, it seems that speech segments with F0 contour manipulations can still clearly be discriminated from the filler noise, a mechanism involved in top-down repair of speech. In contrast, in speech on speech scenarios, competing talker discrimination relies on F0 contours to help across-frequency grouping of the different sound sources (target and maskers). Presumably, this simultaneous grouping is more affected by F0 contour manipulations as suggested by the difference of results observed between the present study and the studies from Binns and Culling (2007), Meister et al.(2011), and Miller et al. (2010).

Even if interrupted speech intelligibility did not significantly differ between F0 contour manipulations (except for the inverted contours), it is still possible that our participants did require more effort to perform the task with the atypical F0 contours. Moreover, speech redundancy is present a different layers of speech processing (Assmann and Summerfield, 2004), and other cues, instead of F0 contours, might be used for prosody processing, such as duration and intensity. Depending on the task difficulty (affected by the amount of information in the speech stimuli), listeners may rely differently on prosodic information. For example, Cutler and Darwin (1981) showed that sentence stress (which contributes to better intelligibility via better lexical segmentation) can be predicted even without F0 contours for continuous speech in silent. Moreover, for connected speech, Wingfield et al. (1984) observed an advantage for report of speech read as list of words (without prosody, i.e. same stress, and same intensity and pitch variations for every word, same vowel duration, same pause between words, but also with better articulation of each word) over normal prosody at normal speech rate. However, this advantage disappeared at higher speech rate (time-compressed speech). One possible suggestion is that the redundancy added to speech from normal prosody is relevant when the task becomes harder by reducing the processing time (i.e. increasing the cognitive load), especially using duration cues in prosody (a deficit of flat F0 over normal F0 being only observed at normal speech rate in Wingfield et al., 1984). This is in line with the fact that other cues, instead of F0 contours, might be used for prosody processing. Indeed, intensity and duration are good indicators of prosodic information as they covary with F0 contours and provide redundant information for prosody processing. As a result, even when F0 contours are manipulated, intensity and duration can be used for stress perception (Mattys, 2000; Vroomen et al., 1998), segmentation (Spritzer 2007), and intonation recognition (Chatterjee and Peng, 2008; Morrow and Liu, 2013; Peng et al., 2012). For example, Peng et al. (2012) showed that normal-hearing English listeners primarily use the F0 contours to judge whether a sentence is a statement or a question, whereas when listening to vocoded speech, reliance on intensity and duration cues increased because F0 cues were degraded. Peng et al. (2012) also showed that actual cochlear implant (CI) users relied primarily on intensity cues, combined with the degraded F0 cues transmitted through their implant, to recognize intonation patterns. This indicates again that depending on the available information in speech (due to signal degradation or hearing impairment), listeners may adapt to what cues to rely on for different listening tasks. In the present study, inverted F0 contour, which provide misleading and distorted cues, did not complement speech redundancy which might explain the lower performance in speech intelligibility for this condition.

To summarize, the present study shows a relatively small effect of F0 contour manipulations on intelligibility of interrupted speech and no effect on phonemic restoration. Confirming Clarke et al. (2014) findings, these results indicate that top-down repair of speech could be robust to atypical voice cues, suggesting that listeners may partly compensate for the degraded voice cues. It is possible that linguistic information, such as the sentential context, which plays an important role in the restoration mechanisms, helped overcome the negative effects of manipulated F0 contours. Another possibility is that participants relied on other prosodic cues, such as intensity and duration, that covary with F0 contours for prosody processing. Presumably, a combination of both mechanisms may occur to achieve best possible performance. Practically, it seems important for CI users, who show difficulties understanding interrupted speech and performing phonemic restoration (Bhargava et al., 2014), to have good linguistic skills, to learn to use other available cues, and/or to have residual hearing preservation to complement speech redundancy of their degraded speech cues at the different levels (acoustic, linguistic). This additional information would strengthen the primordial interaction between the bottom-up cues and top-down mechanisms necessary for perceptual restoration of interrupted speech.

3.5. Conclusion

- Only inverted F0 contour, which provide misleading F0 cues, had a negative effect on interrupted speech intelligibility, presumably because of the distorted clause boundaries, and the impaired accentuation and access to content word.

- Our other F0 contour conditions seemed to still complement speech redundancy to overcome the manipulated F0 cues. No effect of F0 contour manipulations on phonemic restoration were observed, presumably because of the prevalence of the linguistic context and/or the use of other prosodic cues such as intensity and duration.

Acknowledgments

The authors would like to thank Floor Burgerhof and Wilke Bosma for transcribing participant responses, as well as the participants. This study was supported by a VIDI grant from the Netherlands Organization for Scientific Research, NWO, awarded to Deniz Başkent (grant no. 016.096.397; from Netherlands Organization for Health Research and Development, ZonMw). Further support came from a Rosalind Franklin Fellowship from University of Groningen, University Medical Center Groningen, and funds from Heinsius Houbolt Foundation. The study is part of the 'Healthy Aging and Communication' research program of the Otorhinolaryngology Department of University Medical Center Groningen.